CLAIMS

We claim:

1.   An analysis method for large and/or complex biological data sets from molecular biology experiments, the method comprising:

      a)   importing data in a table data structure;

      b)   comparing data points;

      c)   calculating an optimized data representation; and

      c)   displaying the representation

2.   Method according to claim 1 where for the analysis of large and/or complex biological data sets from arrayed biomolecules or derivatives/substitutes.

3.   Method according to claim 2 whereby Step a) is modifying the original measurements to account for the experimental design and/or to emphasize aspects in the following analysis steps by using one of or combinations of methods listed below:

Shifting of values

Scaling of values

Exclusion of outlets

Merging of multiple measurements

Correction of neighbor influences

Selection of characteristics subsets.

4.   Method according to claim 2 whereby Step b) is performed by eigenvalue calculation on a generalized similarity table for the data points and the derived eigenvectors of the similarity table define the optimized representation.

5.   Method according to claim 2 whereby Step b) is performed on a generalized similarity table for the descriptive variables.

6.   Method according to claim 2 whereby Step b) is performed by the analysis or a generalized similarity table for the descriptive variables to derive the representation of the data points.

7.   Method according to claim 2 whereby Step b) can comprise one or more of the methods in claim 4 and where the similarity matrix used is chosen to be of smaller dimensions.

8.   Method according to claim 7 whereby representation of the data points may be calculated from a similarity table for the descriptive variables by the following procedures:

a)   Eighenvalues and eigenvectors are derived from the similarity table of descriptive variables; and

b)   The eigenvectors and the representation of data points is subsequently constructed from the eigenvectors of descriptive variables as linear combinations of the original data with linear factors taken from the eigenvectors of the descriptive variables.

9.   Method according to claim 2 whereby Step c) may consist of one of several of the following procedures:

a)   graphical visualization of the optimized data representation;

b)   graphical visualization of descriptive variables ; and

c)   graphical visualization of both, data points and descriptive variables in a common representation to highlight relationships.

10.   Method according to claim 7 whereby graphical visualization of the optimized data representation is performed by placing data points at the co-ordinates obtained in the eigenvectors.

11.   Method according to claim 8 whereby the optimal subset of eigenvectors is chosen for maximal explanation of the variance in the data as indicated by the biggest corresponding eigenvalues.

12.   Method according to claim 7 whereby graphical visualization of the descriptive variables is performed by placing descriptive variables at the co-ordinates obtained in the eigenvectors.

13.   Method according to claim 7 whereby graphical visualization for both, data points and descriptive variables in common representation is performed by placing descriptive variables and data points at the co-ordinates obtained in the corresponding eigenvectors and where the representation of variable may be scaled independently.

14.   Method according to claim 7 whereby external biological or chemical or medical information is imported into the representation according to co-ordinates as calculated by a projection onto the eighenvectors and where their co-ordinates may be scaled independently.

15.   Method for the analysis of large and/or complex biological data sets by extracting inherent structures of optimized explanatory power.

16.   Method according to claim 1 whereby the analysis is performed on a computer.

17.   An analysis method for large and/or complex biological data sets from molecular biology experiments, the method comprising:

a)   importing data obtained from the experiments;

b)   calculating a data representation using an algorithm; and

c)   displaying the data representation in a display format wherein the metabolic path may be observed.

18.   An analysis method for large and/or complex biological data sets from molecular biology experiments, the method comprising:

a)   importing data obtained from the experiments;

b)   calculating a data representation using an algorithm;

c)   displaying the data representation;

d)   selecting a data point of interest; and

e)   displaying the selected data point so that the selected data point can be distinguished from non-selected data points.

19.   The method of claim 18, wherein the selected data point represents a selected gene, the method further comprising:

f)   displaying data points representing the selected gene so that the data points are displayed as selected data points.

20.   An analysis method for large and/or complex biological data sets from molecular biology experiments, the method comprising:

   a)   importing data obtained from the experiments;

   b)   calculating a data representation using an algorithm;

   c)   displaying the data representation;

   d)   selecting a data point of interest; and

   e)   using a data base search to obtain additional information regarding the compound, gene, cell, virus, sequence, or substance represented by the selected data point.

21.   A computer implemented method for analysis of large and/or complex biological data sets from molecular biology experiments, the method comprising:

   a)   importing data obtained from the experiments into a computer data storage system;

   b)   calculating a data representation of at least a portion of the imported data using a computer implemented algorithm; and

   c)   displaying the data representation on a computer display in a display format wherein the metabolic path may be observed.

22. A computer readable medium having computer readable program code for analysis of large and/or complex biological data sets from molecular biology experiments, the computer readable medium and a computer input/output system being capable of working together to carry out the steps of:

a)   importing data obtained from the experiments into a computer data storage system;

b)   calculating a data representation of the imported data using a computer implemented algorithm; and

c)   displaying the data representation on a computer display in a display format wherein the metabolic path may be observed.

23.   A computer implemented method for analysis of large and/or complex biological data sets from molecular biology experiments, the method comprising:

a)   importing data obtained from the experiments into a computer data storage system;

b)   calculating a data representation using a computer implemented algorithm;

c)   displaying the data representation on a computer display;

d)   selecting a data point of interest from the displayed data; and

e)   displaying the selected data point so that the selected data point can be distinguished from non-selected data points on the computer display.

24.   A computer readable medium having computer readable program code for analysis of large and/or complex biological data sets from molecular biology experiments, the computer readable medium and a computer input/output system being capable of working together to carry out the steps of:

a)   importing data obtained from the experiments into a computer data storage system;

b)   calculating a data representation using a computer implemented algorithm;

c)   displaying the data representation on a computer display;

d)   selecting a data point of interest from the displayed data; and

e)   displaying the selected data point so that the selected data point can be distinguished from non-selected data points on the computer display.

25.   A computer implemented method for analysis of large and/or complex biological data sets from molecular biology experiments, the method comprising:

a)   importing data obtained from the experiments into a computer data storage system;

b)   calculating a data representation using a computer implemented algorithm;

c)   displaying the data representation on a computer display;

d)   selecting a data point of interest from the displayed data; and

e)   using a computer data base search engine to obtain additional information regarding the compound, gene, cell, virus, sequence, or substance represented by the selected data point.

26.   A computer readable medium having computer readable program code for analysis of large and/or complex biological data sets from molecular biology experiments, the computer readable medium and a computer input/output system being capable of working together to carry out the steps of:

a)   importing data obtained from the experiments into a computer data storage system;

b)   calculating a data representation using a computer implemented algorithm;

c)   displaying the data representation on a computer display;

d)   selecting a data point of interest from the displayed data; and

e)   using a computer data base search engine to obtain additional information regarding the compound, gene, cell, virus, sequence, or substance represented by the selected data point.